Credible equilibria in non-finite games and in games without perfect recall

P. Corcho Sánchez[a] and J.L. Ferreira García[b]

[a]*Universidad de Extremadura, Facultad de CC. EE. y Empresariales, Departamento de Economía Aplicada y Organización de Empresas, E-06071 Badajoz, Spain*

E-mail: pcorcho@unex.es

[b]*Universidad Carlos III de Madrid, Facultad de CC. EE. y Jurídicas, Departamento de Economía, E-28903 Getafe (Madrid), Spain*

E-mail: jlferr@econ.uc3m.es

Credible equilibria were defined in Ferreira et al. [6] to handle situations of preferences changing along time in a model given by an extensive form game. This paper extends the definition to the case of infinite games and, more important, to games with non-perfect recall. These games are of great interest in possible applications of the model, but the original definition was not applicable to them. The difficulties of this extension are solved by using some ideas in the literature of abstract systems and by proposing new ones that may prove useful in more general settings.

Keywords: credible, infinite games, imperfect recall, semistable partitions, stable sets, ugly sets

1. Introduction

Situations in which players' preferences may change over time have been studied by many authors. A partial list of relevant works includes Strotz [16], Pollak [13–15], Phelps and Pollak [12], Von Weizsaecker [18], Blackorby et al. [4] and Hammond [9].

One consequence of introducing the possibility of changing preferences is that decisions may not be consistent in the classical sense: decision makers may regret having made certain choices in the past, even in situations of perfect information. Ferreira et al. [6] propose to model these situations as extensive form games where the different information sets of a given player define a set of agents of that player. They consider only finite games and, after endowing each agent with a utility function of his own, define recursively an equilibrium concept that takes into account two important features. First the equilibrium has to be immune to the possibility of coalitional deviations by agents of one player and, second, it has to satisfy a certain time consistency, which requires that agents that appear earlier in time are considered first and that further deviations cannot include agents that play earlier than the ones in the first deviation. Agreements satisfying this requirement are called credible equilibria.

For many applications suggested in Ferreira et al., the assumption of perfect recall may not be the natural one. In many game models, a player is not an individual. It can be a state, a political party or any other organization. In such cases, it is natural not only to assume that the different agents within this organization may have different priorities, but to allow the possibility of imperfect recall, i.e., situations in which a given agent does not know some actions taken in the past by other agents. In this paper, we take up the challenge of providing an appropriate extension of credible equilibria to these situations and to the case of an infinite number of players and strategies.

There are many difficulties for such an extension. First, the definition cannot be recursive because we no longer have a finite set of agents. Second, the lack of perfect recall causes that an agent A may not know whether he is playing before or after agent B, and at the same time, agent B may not know if she plays before or after A. This again generates circularities in the definition and demands a clear discussion on the notion of time consistency.

We address the problems of circularity by using ideas from the literature related to the Theory of Social Situations, initiated by Greenberg [8]. One characteristic of this theory is that, using a generalization of the stable sets defined by von Neumann and Morgenstern [17], recursive definitions may be extended to infinite sets. The idea is to divide the set of agreements into a stable partition of two sets, the good and the bad, so that no element in the good set is dominated by any other good element, and every element in the bad set is dominated by some good element. Of course, in our case, the domination has to do with deviations by coalitions of agents of the same player in some appropriate way.

Such a division may not be possible, as in the cases in which one has an infinite space of strategies or a cyclical dominance relation. Our model has both. The second best is then to find a weaker division on the set of agreements. A semistable partition consists of three sets, the good, the bad and the ugly (Kahn and Mookherjee [10]), where the good elements are dominated only by the bad ones, the bad ones are dominated by the good ones and the ugly set is the complement of the union of the good and the bad. This partition is always possible, and it is customary to choose the good and the ugly sets to define the equilibrium.

However, we find that, at least in our case, some ugly elements may be uglier than others. Since the ugly set may be very large (we provide an example with empty good and bad sets), it seems only natural to explore the structure of this set. We do this and propose a new relation on the ugly set based on classes of equivalence and a domination relation on the quotient set. The idea is to identify elements within a cycle. The partition on this ugly set may again not be a stable partition, but we show that the only cause for the existence of such an ugly–ugly set is if the quotient set of the equivalence relation is infinite, since the new dominance relation is acyclical. Thus, to accept this set is not as problematic as to accept the whole original ugly set. The new definition then uses the original good set, the good–ugly and the ugly–ugly, and gets rid of the bad–ugly and the original bad.

It is important to notice that, at this stage, credible equilibria are just a generalization of the basic concept of Nash equilibria. Refinements like sequential, perfect and the like are still to be adjusted to the framework of changing preferences.

The paper is divided as follows. Section 2 is a brief presentation of the original model in which credible equilibria are defined. Section 3 extends the definition of credible equilibria to infinite games, but still assumes perfect recall. Section 4 drops the perfect recall assumption and studies some of the difficulties that arise by so doing. Preliminary versions of credible equilibria are provided. Section 5 studies the ugly set and proposes a partition on this set. A final version of credible equilibrium is provided. The techniques in this section may be useful for many other definitions within the theory of social situations. Section 6 discusses existence and section 7 concludes.

2. Credible equilibria in finite games

The concept of credible equilibrium was introduced by Ferreira et al. [6] to handle situations in which priorities change during the conflict. In this preliminary section, we present their extensive-form game model and the solution concept, called *credible equilibrium*, which generalizes the concept of Nash equilibrium. This model is that of a finite game with perfect recall. In the following sections, we drop these assumptions and solve a number of conceptual and technical difficulties that arise in the process.

Let $\Gamma = (T, P, U, C, p)$ be a game in extensive-form without payments at the endpoints. Here, T is a tree, $P = \{P_0, P_1, \ldots, P_n\}$ is the players' partition on the non-final nodes of T; $U = (U_0, \ldots, U_N)$, where $U_i = \{u_{i.j}\}_{j=1}^{k_i}$ is the partition of P_i into information sets; $C = \{C(u_{i.j})\}_{i=1,\ldots,n}^{j=1,\ldots,k_i}$ is a correspondence, where $C(u_{i.j})$ is the set of choices which are available to player i at information set $u_{i.j}$; $p = \{p(u_{0j.})\}_{j=1,\ldots,k_0}$ is a vector-valued function, where $p(u_{0j})$ is a probability distribution on chance moves at u_{0j}.

To complete the description of the model, we endow each agent $i.j$ with a von Neumann–Morgenstern utility function $h_{i.j}$, defined on lotteries over endpoints of T. Formally, the game with utilities changing during the play is a six-tuple,

$$\Gamma = (T, P, U, C, p, h),$$

where T, P, U, C, p are as above, and $h = (h_1, \ldots, h_n)$, where, for an endpoint z, $h_i(z) = (h_{i.1}(z), h_{i.2}(z), \ldots, h_{i.k_i}(z))$.

Notice that for each i, we obtained different agents $i.j$ at different information sets $u_{i.j}$ of player i. The interpretation is as follows: at the begining of play, player i believes that he will have the utility function $h_{i.j}$ when he reaches information set $u_{i.j}$.

Throughout this section, we assume that the game is of perfect recall according to the standard definition.

Definition 1. A game form (T, P, U, C, p) is said to be of perfect recall if, for every i ($i = 1, 2, \ldots, n$) and every two information sets $u_{i.j}$ and $u_{i.k}$ of the same player i, if one node y, $y \quad u_{i.k}$, comes after a choice c at $u_{i.j}$, then every node x in $u_{i.k}$ comes after the same choice c.

By Kuhn's theorem [11], in the presence of perfect recall, one can restrict the analysis to behavioral strategies.

Formally, a behavioral strategy $s_{i.j}$ of agent $i.j$ is a probability distribution over the choices $c_{i.j}$ at $u_{i.j}$. We denote by $S_{i.j}$ the set of these strategies. The set of behavioral strategies for player i is $S_i = \times_{j=1,\ldots,k_i} S_{i.j}$. An n-tuple of behavioral strategies is $s = (s_1, \ldots, s_n)$, and the set of these n-tuples is $S = \times_{i=1,\ldots,n} S_i$.

Let Q be a set of agents belonging to the same player. We denote by $-Q$ the set $M \setminus Q$, where M is the set of all agents (not only of the same player). For a strategy s, we denote by s_Q the vector of strategies $(s_{ij})_{ij \quad Q}$. For simplicity, we write (s_{-ij}, s'_{ij}) to express a deviation of agent $i.j$ from s'_{ij}. For s and s' in S, we write $s' \succ_{ij} s$ iff $h_{ij}(s') > h_{ij}(s)$.

Now, we are ready for the definitions:

Definition 2. We say that $i.j$ plays after $i.j_0$ if $i.j = i.j_0$ or if every path from u_{ij} to the root passes through u_{ij_0}.

Definition 3. Let Γ be a game of perfect recall with utilities changing during the play, let s be an n-tuple of behavioral strategies, and let Q *b*e a set of agents of player i, containing an agent $i.j_0$ and possibly some of i's agents that play after $i.j_0$. A strategy s'_Q is said to be a credible deviation from s, struck by agent $i.j_0$ using Q, if

(i) $s' \succ_{ij_0} s$, where $s' = (s'_Q, s_{-Q})$.

(ii) $s' \succ_{ij} (s'_{-ij}, s_{ij})$ for all $i.j \quad Q$, $i.j \neq i.j_0$.

(iii) No agent of i, whether in Q or not, that plays after $i.j_0$ can strike a credible deviation from s'.

Definition 4. Let Γ be a game of perfect recall with utilities changing during the play. A behavioral strategy profile s is called a credible equilibrium (CrE) if no agent can strike a credible deviation from it.

Remark 1. These concepts are defined recursively. For further information concerning these concepts, see Ferreira et al. [6].

3. Extending credible equilibria to infinite games

In this section, we extend the definition of credible equilibria to games with an infinite number of players, that is, $N = \{1, 2, 3, \ldots\}$. For this purpose, we will follow the approach in Asheim [2] where the notion introduced by von Neumann and Morgenstern of abstract stable sets of appropriate abstract systems is used to extend recursive definitions[1)] to the infinite case.

A von Neumann and Morgenstern abstract system (AS) is a pair $(D, \succ)$ where D is an abstract set and $\succ$ is a dominance relation. The notation $f \succ d$ will be interpreted to mean that f dominates d. Let $(D, \succ)$ be an abstract system, and let $f \in D$. The dominion of f, denoted by $\triangle(f)$, is the set of all elements of D that f dominates according to the dominance relation $\succ$:

$$\triangle(f) = \{d \in D: f \succ d\},$$

$$\triangle(F) = \cup\{\triangle(f) : F \subset D\}.$$

That is, an element d in D belongs to $\triangle(F)$ if it is dominated by some element in F. A set $F \subset D$ is a von Neumann and Morgenstern abstract stable set (ASS) for the system $(D, \succ)$ if $F = D \setminus \triangle(F)$.

Let Γ be a game with utilities changing during the play. Inspired by definition 4, an abstract system $(D, \succ)$ is introduced. Let i be a player. For every agent, $i.j$, we define $\mathcal{Q}_{ij}$ to be the set of all agents of i that play after $i.j$:

$$D = \bigcup_{i \in N} \{(i.j, Q_{ij}, s) : i.j \in Q_{ij} \subset \mathcal{Q}_{ij} \quad s \in \times S_i,\ j \in \{1, \ldots, k_i\}\},$$

$$(i.j, Q_{ij}, s) \succ (i.k, Q_{ik}, \hat{s})$$

iff

(i) $i.j$ plays after $i.k$.

(ii) $h_{i.j}(s) > h_{i.j}(\hat{s})$.

(iii) $h_{i.h}(s) > h_{i.h}(\hat{s}_{i.h}(s_{-i.h})$ for all $i.h \in Q_{i.j}$.

(iv) $s_{-Q_{ij}} = \hat{s}_{-Q_{ij}}$.

The next proposition relates the ASS of $(D, \succ)$ to the definition of credible equilibrium for finite games of perfect recall with utilities changing during the play and allows for a definition applicable to infinite games of perfect recall with utilities changing during the play.

Denote by Γ_{ij}^{s} the game obtained from Γ by fixing the strategy of every agent of every player according to s, except $i.j$ and agents of player i that play after $i.j$. Note that Γ_{ij}^{s} is a one-person game, possibly with several agents, i.e., at the beginning of the game, players think that that is the game agent $i.j$ is facing, given that every agent

1) The recursion being on the number of players, pure strategies, periods, etc.

other than himself and his followers play s. Denote by $i.j_0$ an agent of player i that does not play after any other agent of that player.

Proposition 5. Let K be an ASS of $(D, \succ)$. Then, for finite games (finite number of agents and pure strategies) of perfect recall with utilities changing during the play, we have that for all $i.j$, $Q_{ij} \subset \mathcal{Q}_{ij}$ and $s \quad \times_{i=1,\ldots,n} S_i$, $(i.j,\ Q_{ij},\ s) \quad K$ if and only if s is a credible equilibrium in Γ. In particular, $A = \{s : \forall i \quad N$ and $\forall\, i.j_0\ (i.j_0, \mathcal{Q}_{ij_0}, s) \quad K\}$ is the set of credible equilibria of Γ.

Proof. First, if for all $i \quad N$, for all $i.j$ and $Q_{ij} \subset \mathcal{Q}_{ij}$, then $(i.j,\ Q_{ij},\ s) \quad K$ is not dominated by any element in D, and no agent of i that plays after $i.j$, say $i.k$, can strike a credible deviation with $\hat{s} \succ_{ik} s$ and $\hat{s} \succ_{ih} (s_{ih},\ \hat{s}_{-ih})$. This is true for all i, $i.j$ and, in particular, for $i.j \neq i.k$ which, by definition 4, implies that s is a credible equilibrium.

To prove the converse, let s be a credible equilibrium of Γ, so that no agent after $i.j$ can strike a credible deviation. If for some $i.j$ and Q_{ij}, then $(i.j,\ Q_{ij},\ s) \quad K$ is dominated by some element in K. This means that some agent that plays after $i.j$ can strike a credible deviation, i.e. $\quad i.k,\ Q_{ik}$ and $s' = (s'_{Q_{ik}},\ s_{-Q_{ik}})$, such that $(i.k,\ Q_{ik},\ s') \succ (i.j,\ Q_{ij},\ s)$. Since $(i.k,\ Q_{ik},\ s') \quad K$, s' is an equilibrium in $\Gamma^{s'}_{i.k}$ and, hence, s' is a deviation from s struck by $i.k$. If $i.k \neq i.j$, this contradicts the fact that s is a credible equilibrium. If $i.k = i.j$, by theorem 5.1 in Ferreira et al. [6], there exists another deviation s^* struck by $i.j$ such that s^* is a credible deviation from s. But then s is not a credible equilibrium, which is a contradiction. □

The uniqueness of the abstract stable set is established in the following proposition.

Proposition 6. Let F and T be abstract stable sets of the abstract system $(D, \succ)$ associated with a finite game, as defined above; then $F = T$.

Proof. Let $\{(i.j_k,\ Q_{ij_k},\ s_k)\}_k$ be an infinite sequence such that $(i.j_{k+1},\ Q_{ij_{k+1}},\ s_{k+1}) \succ (i.j_k,\ Q_{ij_k},\ s_k)$, for all k. By the finiteness of the set of agents and of the set of strategies and because the dominance relation requires that ij_{k+1} play after ij_k, we have that, in the tail of the sequence, $Q_{ij_k} = Q_{ij_{k+1}}$ and $ij_{k+1} = ij_k$. This tail is transitive. According to corollary 4b in Arce and Kahn [1], this is a sufficient condition for the uniqueness of the abstract stable set if it exists. □

Since the characterization of credible equilibria given in proposition 5 is not based on recursion, it can be used to formulate a general definition of the concept, covering both finite and infinite games.

Definition 7. Consider a game of perfect recall with utilities changing during the play and with an infinite or finite number of players and strategies. A sequence

$s = (s_1, s_2, \ldots)$ is said to be a credible equilibrium of Γ if and only if there is an abstract stable set (ASS), K, for the associated system $(D, \succ)$ such that for all $i \quad N$, for all ij_0 and for all $Q_{ij_0} \subset \mathcal{Q}_{ij_0}$, $(i.j_0, Q_{ij_0}, s) \quad K$.

The abstract set K is interpreted as a social norm, where every point in K is equally reasonable: no one dominates any other in the social norm and any point outside the social norm is dominated by some element inside it.

A major difficulty in this approach is that when the abstract system $(D, \succ)$ is not finite or $\succ$ is cyclical[2)], stable sets may not exist. Kahn and Mookhejee [10] provide examples of games with infinite action spaces where the stable set does not exist.

4. Imperfect recall

In this paper, we follow the standard *ex ante* view on strategy choice. This implies that each agent evaluates the possible outcomes resulting from behavior and mixed strategies before the game.

The perfect recall condition on information partitions is a basic assumption in the study of extensive games. This condition expresses the idea that a player remembers what he did and what he learned, and seems natural for rational players. In our model, imperfect recall means that agents may or may not have the ability to observe moves by other agents of the same player. Since agents play only at one information set, the set of mixed strategies for that agent coincides with the set of behavior strategies with and without the perfect recall assumption. The difference that imperfect recall introduces in our model is on the relation *play after*.

Ferreira et al. [6] suggest some applications in which perfect recall may not be a reasonable requirement. However, they do not pursue this idea further. In this section, we study this extension and propose a notion of credible equilibrium for games without perfect recall. As we will see, one consequence of not having perfect recall is that we can no longer produce a recursive definition. When we try the ASS approach, some difficulties arise since we cannot guarantee the existence of a unique stable set even with a finite number of players. Hence, we use the more general concept of semi-stability.

Another consequence of not assuming perfect recall has to do with the specification of the permitted deviations. According to condition (iii) of definition 3 of the original credible deviations, after a credible deviation no other agent after the one who proposed the deviation should be able to find a new one. Now, this agent is not required to play with positive probability according to the first deviation, yet if he can strike a new one, that deviation was not credible. In the original framework (finiteness and perfect recall), this did not represent any problem, since agents playing with probability zero will never gain from a deviation as they can only instruct agents after

[2)] These cases could be an infinite game or imperfect recall.

them, who, again, play with probability zero. Therefore, their actions will not change the expected payoffs when evaluated at the beginning of the play. Without perfect recall, this does not need to be the case. An agent that plays with probability zero may instruct agents that, according to him, may play after him and that, by changing their behavior, can make him play with positive probability. This is a very unnatural deviation to be permitted. Hence, we have to prevent that possibility in the new definition.

We start by extending the relation *play after* to the case of imperfect recall and by giving an example of how this relation may be cyclical.

Definition 8. Let Γ be a game with or without perfect recall and with utilities changing during the play. We say that $i.j$ plays after $i.k$ according to strategy s if there exist a path from the root of the game passing through $y \quad u_{i.k}$ and $x \quad u_{i.j}$, and if both agents play with positive probability if s is played.

Remark 2. If one uses this new relation *plays after ... according to strategy ...* in definition 3, the definition does not change as it refers to games of perfect recall.

The example in figure 1 illustrates how the lack of perfect recall may make the relation *play after* cyclical. Observe that, according to the strategy in which every agent randomizes between his strategies, agent 1.2 can play after 1.3 and agent 1.3

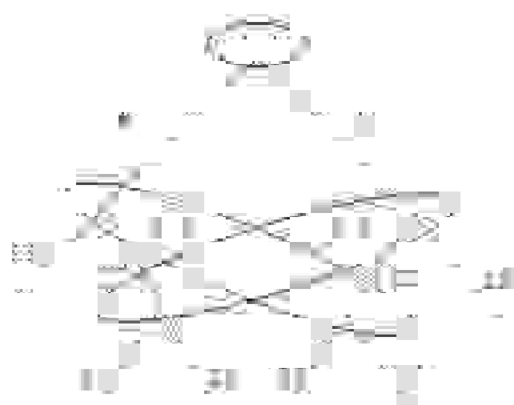

Figure 1.

after 1.2, because their information sets cross each other. Notice also that, according to strategy (R, $L2$, $R3$), agent 1.2 plays with probability zero. If we allowed this player to strike a deviation with agents that may play after him, he could suggest to agent 1.3 to play l.3, thus making him play with positive probability. These are the kind of situations we prevented in the definition.

Now we extend the definition of credible equilibrium to games of imperfect recall, in which coalitions of players communicate prior to actual play and make non-binding agreements on strategy choices. We wish to know which agreements are stable in such environments.

For a strategy s, we denote by $Q_{i\,j}(s)$ the set of all agents of i that play after $i.j$ according to s. Inspired by definition 4 and by the above discussion, an abstract system is introduced:

$$D = \{(i.j, Q, s)\ ;\ i.j \quad Q \subset \mathcal{Q}_{i.j}(s), \quad s \quad \times_i S_i\}, \tag{1}$$

$$(i.j, Q, s) \succ\succ (i.k, H, \hat{s})$$

iff

(i) $i.j$ plays after $i.k$ according to $\hat{s}$.

(ii) $h_{i.j}(s) > h_{i.j}(\hat{s})$.

(iii) $h_{i.h}(s) > h_{i.h}(\hat{s}_{i.h},\ s_{-i.h)})$, for all $i.h \quad Q$.

(iv) $s_Q = \hat{s}_{-Q}$.

An element of D, which will be called an agreement, is a specification of the moves to be taken by all members in the agreement, for a given list of moves for all other players. Condition (i) allows the analysis of games with crossing information sets as in figure 1. Only if agent $i.j$ is reached with positive probability under $\hat{s}$ can he strike a deviation from $\hat{s}$. Condition (ii) states that $i.j$ prefers the proposed deviation s rather than the initial strategy profile $\hat{s}$.Condition (iii) implies that when $i.h$, a member of Q, comes to play, he prefers to comply with the deviation rather than with the strategy $\hat{s}$, given that all other agents in Q follow the deviation s. Note that conditions (ii)–(iii) of the dominance relation impose requirements only on players who play after $i.j$ according to s.

Our objective is to determine whether an n-player agreement is stable, i.e. whether it is never dominated or is dominated only by an agreement that is dominated by an agreement that is never dominated or … and so on.

Von Neumann and Morgenstern established sufficient conditions for the existence and uniqueness of the abstract stable set by considering properties of an acyclical dominance relation on the abstract set. Unfortunately, the dominance relation $\succ\succ$ is not acyclical, therefore we cannot use those sufficient conditions to show existence of the abstract stable set. This situation departs from the case of a finite number of players and strategies and with an acyclical dominance relation, where there always exists a unique partition of the abstract system into stable and non-stable sets.

Kahn and Mookerjee [10] provide an example of a game in which a stable partition does not exist when considering an abstract system based on the concept of coalition-proof equilibrium. The example involves an infinite set of strategies. For this reason, these authors make a modification on the definition of stability. They show that a weaker version of stable partitions, called semistable partitions, always exists for arbitrary sets of objects and arbitrary dominance relations.

Definition 9. Let $(\mathcal{A}, >>)$ be an abstract system. A trio of subsets $\{\mathcal{G}, \mathcal{B}, \mathcal{U}\}$ of $\mathcal{A}$ form a semistable partition (SSP) of $\mathcal{A}$ if:

(1) $\mathcal{B}$ consists of all elements dominated by elements in $\mathcal{G}$, i.e., $x \in \mathcal{B}$ iff $\exists x' \in \mathcal{G}$ such that $x' >> x$.

(2) $\mathcal{G}$ consists of all elements which, if they are dominated by some element, then they are dominated by elements in $\mathcal{B}$, i.e., $x' \in \mathcal{G}$ iff, whenever $x >> x'$, then $x \in \mathcal{B}$.

(3) $\mathcal{G} \cap \mathcal{B} = \emptyset$ and $\mathcal{U} = \mathcal{A} \setminus (\mathcal{G} \cup \mathcal{B})$.

$\mathcal{G}$ is called the good set, $\mathcal{B}$ the bad set and $\mathcal{U}$ the ugly set.

Ideally, we would like to have a unique stable partition (i.e. $\mathcal{U} = \emptyset$) because a semistable partition that is not stable contains an ugly set, which makes the solution concept ambiguous. Unfortunately, in our case, the abstract system $(D, \succ\succ)$ does not admit a $(\succ\succ)$-stable partition in general, as will be seen later on when studying the example in figure 3. A semistable partition is a weaker concept than a stable partition in the sense that the good and the bad sets do not exhaust the set of all agreements. One result on a semistable partition (Arce and Kahn [1]) is that the ugly set contains cycles or infinite sequences of agreements dominating one another, none of which is dominated by a good agreement. We will study the ugly set of $(D, \succ\succ)$ in more detail in the next sections. Before doing this, notice that semistable partitions may not be unique. Our extensions of credible equilibria will be based on one interesting class of them: The minimal semistable partition.

Definition 10. A $(\succ\succ)$-semistable partition $\{G^*(\succ\succ), B^*(\succ\succ), U^*(\succ\succ)\}$ of D is minimal if it satisfies $G^*(\succ\succ) \subset G(\succ\succ)$, $B^*(\succ\succ) \subset B(\succ\succ)$ and $U^*(\succ\succ)$ is the complement of $G^* \cup B^*$ for every semistable partition $\{G(\succ\succ), B(\succ\succ), U(\succ\succ)\}$ of D. These sets $\{G^*, B^*, U^*\}$ are called strictly good, strictly bad and strictly ugly, respectively.

Minimal semistable partition can be constructed as follows (Kahn and Mookherjee [10]). First define the sets G_0^* and B_0^*:

$$G_0 = \{(ij, Q, s) \in D : \text{ not } \exists (ih, H, \hat{s}) \in D : (h, H, \hat{s}) \succ\succ (ij, Q, s)\},$$

$$B_0 = \{(ij, Q, s) \in D : \exists (h, H, \hat{s}) \in G_0^* : (ih, H, \hat{s}) \succ\succ (ij, Q, s)\}.$$

Next, inductively define G_k^*, B_k^* with $k = 1\ldots$,

$$G_k^* = \{(ij, Q, s) \in D : \text{ if } (ih, H, \hat{s}) \succ\succ (ij, Q, s) \text{ then } (ih, H, \hat{s}) \in B_{k-1}^*\},$$

$$B_k^* = \{(ij, Q, s) \in D : \exists (h, H, \hat{s}) \in G_k^* : (h, H, \hat{s}) \succ\succ (ij, Q, s)\}.$$

Define $G^* = \bigcup_{k=0}^{\infty} G_k^*$ and $B^* = \bigcup_{k=0}^{\infty} B_k^*$. And finally, define U^* as the complement of $G^* \cup B^*$ in D.

Once the semistable partition is defined, one can use the elements on the strictly good set to define the equilibrium. The possible existence of a non-empty ugly set opens the possibility of a stronger and weaker version of equilibrium.

Definition 11. Let Γ be a game of imperfect recall as defined above, with a finite or infinite number of players and strategies, let $(D, \succ\succ)$ be an abstract system as defined in (1), and let $\{G^*, B^*, U^*\}$ be a minimal semistable partition of D. A strategy s is said to be a strongly-credible equilibrium of Γ if for all $i.j_0$ and $Q_{ij_0} \subset \mathcal{Q}_{ij_0}$, (ij_0, Q_{ij_0}, s) $G^*(\succ\succ)$.

Our solution concept says that (ij_0, Q_{ij_0}, s) is stable if it is dominated only by elements belonging to the strictly bad set. A weaker notion could also be defined as follows:

Definition 12. Let Γ, $(D, \succ\succ)$ and $\{G^*, B^*, U^*\}$ be defined as before. A strategy s is said to be a weakly-credible equilibrium of Γ if for all $i.j_0$ and $Q_{ij_0} \subset \mathcal{Q}_{ij_0}$, (ij_0, Q_{ij_0}, s) $G^*(\succ\succ) \cup U^*(\succ\succ)$.

The weaker solution says that an element of D is stable if it is not dominated by any element in the strictly good set. These are standard ways of defining equilibria in the literature of semistability; however, none of these definitions is entirely satisfactory, first because the set G^* may be empty and, second, because some ugly elements may be uglier than others. The following section considers these points.

We end this section with an example of a game of imperfect recall in which the dominance relation turns out to be acyclical. The example illustrates the different implications of applying the original definition of credible equilibria and the ones just proposed.

Figure 2.

Consider the game in figure 2; $(L1, L2)$ is not credible according to the original definition because agent 1.1 can instruct himself and agent 1.2 to move $(R1, R2)$. If agent 1.2 knew that agent 1.1 was moving right, he would certainly have chosen to comply. But he does not know, and if he complies, then it behooves agent 1.1 to remain at $L1$ so as to get 3. So, the deviation ($R1$, $R2$) is perhaps not safe.

We could conclude that the original definition of credible equilibria is not appropriate in this context, and certainly it was never intended to be. If we apply definition 2 or 3, we get the following chain of dominations:

$$(1.1, N, (L1, L2)) \prec\prec (1.1, N, (R1, R2)) \prec\prec (1.1, \{1.1\}, (L1, R2))$$
$$\prec\prec (1.2, \{1.2\}, (L1, L2)).$$

Notice that the last agreement is not dominated (it would need the participation of 1.1 to go ($R1$, $R2$), but this is not possible as 1.1 does not play after 1.2).

Hence, ((1.1, N, ($L1$, $L2$)), (1.1, {1.1}, ($L1$, $R2$)) are in the bad set and (1.1, N, ($R1$, $R2$)), (1.2, {1.2}, ($L1$, $L2$)) are in the good set.

5. Cycles and equivalence classes

Consider the game in figure 3 and, for simplicity, consider only pure strategies. The following relations can be easily checked. The agreements (1.2, N, ($R2$, $L3$)) and (1.3, N, ($L2$, $R3$)) form a cycle since they dominate each other. The agreement (1.3, N, ($R2$, $R3$)) (with $0 < E < 1/3$) is dominated by the two agreements in the cycle

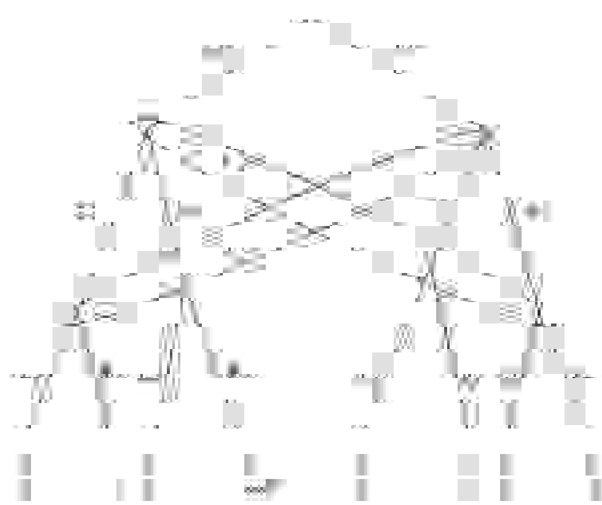

Figure 3.

before, and dominates the agreement (1.2, N, ($L2$, $L3$)). Finally, (1.2, N, ($L2$, $L3$)) does not dominate any other agreement and, therefore, is an element of the ugly set. This suggests the division of the ugly set into a new partition where the elements in a cycle that are not dominated by elements outside the cycle may be considered good. In this section, we present such a division.

One result on the structure of ugly sets that will prove useful is the following (Arce and Kahn [1]): Let $(D, \succ)$ be an abstract system, and $\{G, B, U\}$ a semistable partition on it; then, if $U \neq$, U contains a stack, i.e. an infinite sequence $\{x_i\}_i$ such that $x_{i+1} > x_i$ for all i. Sometimes U contains cycles, which are stacks with a finite range. The strategy in this section consists in identifying the elements within a cycle and in defining a new dominance relation on these.

Definition 13. Let $(A, \succ)$ be an abstract system; a cycle in $(A, \succ)$ is a finite sequence $(x_i)_{i=1,\ldots,n}$ with x_i D that satisfies

$$x_1 \succ x_2 \succ \cdots \succ x_1.$$

Let $\{G, B, U\}$ be a minimal semistable partition associated to $(D, \succ\succ)$. Suppose that U is a non-empty set and that it contains, at least, one cycle. As we saw in the previous example, in this set some ugly elements *may be uglier than others*. We will use the methodology of stable sets to characterize these good–ugly elements in infinite games without perfect recall. Therefore, the idea is to partition the ugly set into two subsets, which can be denoted as the good–ugly and the bad–ugly, with the properties of internal stability and external stability.

From the initial binary relation $\succ\succ$, we define the following binary equivalence relation:

$$(i.j, Q, s) \sim (i.h, H, \hat{s})$$

iff either $(i.j, Q, s) = (i.h, H, \hat{s})$ or there exist sequences $\{(x_i)_i\}$, $\{(y_i)_i\}$ in D such that

$$(i.j, Q, s) \succ\succ x_1 \succ\succ \cdots x_p \succ\succ (i.h, H, \hat{s}) \succ\succ y_1 \succ\succ \cdots y_k \succ\succ (i.j, Q, s).$$

Proposition 14. The binary relation $\sim$ on U defined above is an equivalence relation.

Proof. It is trivial to show that the relation $\sim$ is reflexive and symmetric.

To show transitivity, suppose that $(i.j, Q, s) \sim (i.h, H, \hat{s})$ and $(i.h, H, \hat{s}) \sim (i.k, K, s')$. This implies that either $(i.j, Q, s) = (i.h, H, \hat{s}) = (i.k, K, s')$ or

$$(i.j, Q, s) \succ\succ x_1 \succ\succ \cdots x_p \succ\succ (i.h, H, \hat{s}) \succ\succ y_1 \succ\succ \cdots y_k \succ\succ (i.j, Q, s)$$

and

$$(i.h, H, \hat{s}) \succ\succ \hat{x}_1 \succ\succ \cdots \hat{x}_p \succ\succ (i.k, K, s') \succ\succ \hat{y}_1 \succ\succ \cdots \hat{y}_k \succ\succ (i.h, H, \hat{s}).$$

Then, $(i.j, Q, s) \succ\succ x_1 \succ\succ \cdots (i.h, H, \hat{s}) \succ\succ \hat{x}_1 \cdots \succ\succ (i.k, K, s')$. But this implies that $(i.j, Q, s) \sim (i.k, K, s')$. $\square$

We now consider the quotient set obtained after the equivalence classes of $\sim$ and define a domination relation on it.

Definition 15. The classes of an equivalence relation $\equiv$ on $\mathcal{A}$ are the collections $[a] = \{x \quad \mathcal{A} \setminus x \equiv a\}$. The collection $\{[a]\}_{a \quad \mathcal{A}}$ is denoted by $\mathcal{A}/\equiv$ and called the quotient set[3] of $\mathcal{A}$ by $\equiv$.

From the ugly set $\mathcal{U}$ and the initial relation $\succ\succ$, $(U, \succ\succ)$, we define a new abstract system,

$$((U/\sim), \gg),$$

where the abstract set is the quotient set of equivalence relation defined on the ugly set, and the dominance relation is defined as follows: for $[a], [b] \quad (U/\sim)$,

$$[a] \gg [b]$$

iff $[a] \neq [b]$ and there exist $x \quad [a]$, $y \quad [b]$ and a sequence $\{(x_i)_i\}$ with $x_i \quad U$ such that $x \succ\succ x_1 \succ\succ \cdots \succ\succ x_p \succ\succ y$.

For the general case, a stable partition on $((U/\sim), \gg)$ may not exist. Consider then a minimal semistable partition $\{G_U, B_U, U_U\}$, and the following sets:

$$G_U^*(\sim) = \{x \quad U : [x] \quad G_U(\gg)\} \text{ and } U_U^*(\sim) = \{x \quad U : [x] \quad U_U(\gg)\},$$

With these elements, we can formulate a new extension of credible equilibrium that is more satisfactory than definitions 2 and 3.

Definition 16. Let Γ be a game without perfect recall, with utilities changing during the play and with a finite or infinite numbers of players. A strategy s is said to be a credible equilibrium of Γ if for all $i.j_0 \quad N$ and $Q_{ij_0} \subset \mathcal{Q}_{ij_0}$, $(i.j_0, Q_{ij_0}, s) \quad G(\succ\succ) \cup G_U^*(\sim) \cup U_U^*(\sim)$.

At this point, it seems that one could again consider the possibility that some elements in U_U^* are uglier than others and repeat the whole procedure for this set. The good news is that this is not the case, since the set U_U^* does not contain cycles. This is established in the next proposition.

Proposition 17. Let $((U/\sim), \gg)$ be as defined above; then $\gg$ is acyclical on $(U/\sim)$.

Proof. We use the following result: If a relation is asymmetric, irreflexive and transitive, then it is acyclical. It is trivial to show that the relation $\gg$ is irreflexive and

[3] The reader is referred to Potter [19, chap. 2] for a detailed study on equivalence relations and quotient sets.

asymmetric. To show transitivity, suppose that $[a], [b], [c] \in (U/\sim)$, with $[a] \gg [b]$ and $[b] \gg [c]$. These classes are different and there exist $x \in [a]$, $y, y' \in [b]$ and $z \in [c]$ such that

$$x \succ\succ x_1 \succ\succ \cdots \succ\succ x_p \succ\succ y,$$

$$y' \succ\succ y_1' \succ\succ \cdots \succ\succ y_p' \succ\succ z.$$

Since either $y = y'$ *or* $y \succ\succ y_1 \cdots \succ\succ y'$ (because $y, y' \in [b]$), we have $x \succ\succ \cdots \succ\succ z$, which implies that $[a] \gg [c]$. □

If the abstract set is finite and the binary relation is acyclical, then there exists a unique stable partition. Since we have eliminated the possibility of cycles, the only reason not to have a stable partition is if $(U/\sim)$ is not finite. One interesting question for future research is to find conditions that guarantee the existence of a finite quotient set $(U/\sim)$. It is clear that a finite ugly set generates a finite quotient set, but there should be infinite ugly sets whose quotient sets are finite.

6. Existence

Not only in this work, but in all the literature using abstract stable sets, existence theorems are very difficult to show, even in cases where counterexamples are also hard to find. This seems the price to pay for being able to make non-recursive definitions. This has not been a handicap for this methodology to be fruitful, providing important insights into many aspects of hard problems in game theory, especially those related to the stability of coalitional deviations.

Nevertheless, we can show the existence of an approximation of weakly credible equilibria for finite games with imperfect recall. The approximation is on the line of ε-equilibria. Interestingly enough, in a very different setting, but still within the methodology of abstract stable sets, and with a different kind of proof, Greenberg et al. [20] were also able to prove only the existence of an ε-conservative equilibrium.

Definition 18. Define $(D, \succ\succ_\varepsilon)$ as $(D, \succ\succ)$ except that conditions (ii) and (iii) are replaced with

(ii′) $h_{i.j}(s) > h_{i.j}(\hat{s}) - \varepsilon$.

(iii′) $h_{i.h}(s) > h_{i.h}(\hat{s}_{i.h}, s_{-i.h}) - \varepsilon$, for all $i.h$ in $Q_{i.j}$.

Then define ε-weakly credible equilibria as weakly credible equilibria in definition 12, replacing the dominance relation $\succ\succ$ with $\succ\succ_\varepsilon$. In words, an ε-weakly credible equilibrium is like the weakly credible except that agents do not deviate unless they can win more than $\varepsilon > 0$. The next proposition shows that this new equilibrium exists for an important class of games.

Proposition 19. There exist ε-weakly credible equilibria.

The proof of this proposition is established as a corollary of proposition 21. But before that, we need the following definition.

Definition 20. Define $\Gamma(\delta) = (T, P, U, C, p, h, \delta)$ as $\Gamma = (T, P, U, C, p, h)$, with the only difference being that in $\Gamma(\delta)$, for every agent of every player, all pure strategies must be chosen with a positive probability that is a multiple of $\delta > 0$.

Remark 3. For this definition to apply, it has to be the case that for every agent, the number of strategies divided by δ is an integer. It is always possible to define $\Gamma(\delta)$ games when the original game is finite. Within this new class of games, we can apply the different definitions of credible equilibria. Also notice that, for these games, if one player plays after another according to a strategy, he also plays after him according to any other strategy, so there is no need to make reference to the strategy.

Proposition 21. Let $\Gamma(\delta)$ be a game as defined above. If it is finite then it has at least one weakly credible equilibrium.

Proof. Start with a strategy s. If it is not a weakly credible equilibrium, then there exist $(i.j_0, Q_{i.j_0}, s)$ and $(i.j, Q_{i.j}, s')$ such that $(i.j, Q_{i.j}, s') \succ\succ (i.j_0, Q_{i.j_0}, s)$.

Now, if s' is not a weakly credible equilibrium, then we find a new dominating deviation and so on. If at some point we reach a weakly credible equilibrium, we are done. If not, we can construct a sequence of deviations. Since the number of agents and strategies are finite, we must have the following cycle starting with some agent $i.h$:

$$(i.h, Q_{i.h}, s'') \succ\succ (i.k, Q_{i.k}, s''') \succ\succ \cdots \succ\succ (i.h, Q_{i.h}, s'') \succ\succ \cdots$$
$$\succ\succ (i.j, Q_{i.j}, s') \succ\succ (i.j_0, Q_{i.j_0}, s).$$

There are two possibilities. Either the elements in this cycle belong alternatively to the good and the bad sets, or all of them belong to the ugly set. Consider, then, the first possibility and the case in which $(i.h, Q_{i.h}, s'')$ is in the good set. Since all strategies are completely mixed, all agents in the cycle play after each other according to any strategy (recall the above remark). This means that if s'' is not a weakly credible equilibrium, the only deviations that upset it are due to other players or to agents of i that do not play after $i.h$. Fix the part of the strategy s'' for all the agents of i that play after $i.h$, i.e., fix $s''_{Q_{i.h}}$. With the existing deviation, we can repeat the same process that we started with $(i.j_0, Q_{i.j_0}, s)$. Every time we do this, either we encounter a weakly credible equilibrium or else we fix a strategy for a new set of agents. In the latter situation, by the finiteness of the game, we have eventually fixed a strategy for every agent. The strategy profile that is the result of these fixed strategies must be a weakly credible equilibrium by construction (in fact, it would be a strongly credible equilibrium, since it would belong to the good set) since no agent can find a deviation that is not in the bad set. If $(i.h, Q_{i.h}, s'')$ is in the bad set, $(i.k, Q_{i.k}, s''')$ is in the good set and we can repeat the process to fix $s'''_{Q_{i.k}}$. If the elements in the cycle are in the ugly set, the process can be repeated directly with $(i.h, Q_{i.h}, s'')$ and the final strategy profile would be either in the good or in the ugly set, thus being a weakly credible equilibrium. □

Proof of proposition 19. Since payoffs are a continuous function of the probabilities with which mixed strategies are chosen, for every ε there exists a δ small enough so that every payoff that is the result of a strategy combination in Γ is within the ε-neighborhood of a payoff in Γ(δ). Hence, the weakly credible equilibrium in Γ(δ) is an ε-weakly equilibrium in Γ. □

Remark 4. This proof works with any positive ε, but not with ε = 0. The reason is twofold. First, as $\varepsilon \to \infty$, the number of elements in the cycle also goes to infinity and one cannot apply the argument and, second, the set of credible equilibria need not be compact (see Ferreira et al. for an example).

We end the discussion of existence with an interesting observation. In the original paper by Ferreira et al., the existence of credible equilibria was established as a consequence of having as a subset the set of agent-perfect equilibria, which is non-empty. When the assumption of perfect recall is dropped, in addition to all the difficulties we address in this work, this relation does not hold in general. Figure 4 depicts a counter-example.

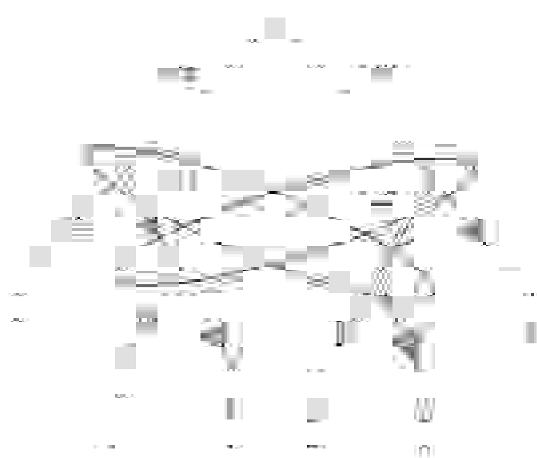

Figure 4.

One can easily check that both agents moving right (dark arrows) is an agent-perfect equilibrium. The same can be said if both agents move left (white arrows). However, the first cannot be a credible equilibrium, since either agent can strike a credible deviation by suggesting to go left.

7. Conclusions

We have extendend the definition of credible equilibria to infinite games and to games with non-perfect recall. By doing so, we encounter a number of difficulties. The possibility of an infinite set of agreements and the fact that players may play one after the other in a cycle called for a non-recursive definition based on semistable partitions. By studying the nature of the ugly set of this semistable partition, we were able to propose a new division of the agreements that is more satisfactory than the standard in the literature.

Other problems remain open. It is important to find necessary conditions to guarantee a finite quotient set in the equivalence relation defined on the ugly set. Other topics for further research include the extension of alternative definitions to credible equilibria, such as optimistic credible equilibria and its variants, as defined in Ferreira et al. [6]. Since all these definitions are extensions of the basic concept of Nash equilibrium, and since the model of changing preferences is intrinsically dynamic, it would be of interest to study extensions of refinements of Nash equilibria to this model.

References

[1] M. Arce and C. Kahn, The mathematics of semistability, Mimeo, University of Illinois at Urbana-Champain, 1991.

[2] B. Asheim, Extending renegotation-proofness to infinite horizon games, Games and Economic Behavior (1991)278.

[3] T. Bergstrom, Maximal elements of acyclic relations on compact sets, Journal Economic Theory 10(1975)413.

[4] C. Blackorby, D. Niessen, D. Primont and R. Russell, Consistent intertemporal decision-making, Rev. Econ. Stud. 40(1973).

[5] B. Chakravorti and C. Kahn, Universal coalition proof equilibrium: Concepts and applications, Bellcore Economics Discussion Paper No. 98, 1993.

[6] J.L. Ferreira, I. Gilboa and M. Maschler, Credible equilibria in games with utilities changing during the play, Games and Economic Behavior 10(1995)284.

[7] J. Greenberg, Deriving strong and coalitional-proof Nash equilibria from an abstract system, Journal of Economic Theory 49(1989)195.

[8] J. Greenberg, *The Theory of Social Situations. An Alternative Game-Theoretic Approach*, Cambridge University Press, 1990.

[9] P. Hammond, Changing tastes and coherent dynamic choice, Rev. Econ. Stud. 43(1976).

[10] C. Kahn and D. Mookherjee, The good, the bad and the ugly: Coalition-proof equilibrium in infinite games, Games and Economic Behavior 4(1992)101.

[11] H.W. Kuhn, Extensive games and the problem of information, Annals of Mathematics Studies 28(1953)193.

[12] E. Phelps and R. Pollack, On second-best national saving and game-equilibrium growth, Rev. Econ. Stud. 35(1968).
[13] R. Pollack, Consistent planning, Rev. Econ. Stud. 35(1968).
[14] R. Pollack, Habit formation and dynamic demand fuctions, Journal Polit. Econ. 78(1970).
[15] R. Pollack, Habit formation and long-run utility fuctions, Journal Econ. Theory 13(1976).
[16] R.H. Strotz, Myopia and inconsistency in dynamic utility maximization, Rev. Econ. Stud. 23(1956).
[17] Von Neumann and Morgenstern, *The Theory of Games and Economic Behavior*, Princeton University Press, 1947.
[18] C. Von Weizsaecker, Notes on endogeneous changes of tastes, Journal Econ. Theory 4(1971).
[19] M.D. Potter, *Sets: An Introduction*, Oxford Science, Clarendon Press, 1990.
[21] J. Greenberg, D. Monderer and B. Shitovitz, Multi-agent situations, Econometrica 64(1996).